共生还是对立？人类与人工智能的辩证关系

胡晓檬　彭凯平　朱昕语

摘要：本文结合多学科的研究成果，综述了人类如何从利益论、威胁论及辩证关系视角看待人工智能（AI）的复杂态度。本文首先探讨了 AI 如何作为一种技术进步，促进了工作效率、决策质量的提高以及在多个领域内的创新应用，反映了对 AI 的积极评价和期望。随后，本文转向 AI 带来的潜在威胁，包括对个人隐私的侵犯、就业市场的冲击以及社会伦理的挑战。最后，从辩证的角度审视了 AI 技术，强调了在 AI 发展过程中平衡技术创新与社会伦理的重要性。未来的研究应结合人类对 AI 态度的多维性和复杂性，致力于 AI 伦理的探索和人机协作的深化。

关键词：人工智能，对 AI 的矛盾态度，AI 伦理，利益论，威胁论，辩证观

1 引言

在人工智能（AI）日益成为我们生活和工作中不可或缺的一部分之际，关于人类如何看待 AI 的讨论变得尤为重要。这种讨论不仅关乎技术发展的方向，也触及到我们对未来社会的期待与担忧。AI 的发展引发了多种观点，其中利益论、威胁论和辩证关系三种视角尤为突出。这些观点不仅体现了对 AI 技术潜力的不同理解和评估，还反映了对 AI 和人类未来关系的深切考量。从利益论角度看，AI 被视为一种能够极大提升社会生产力和效率的技术。AI 的支持者认为，通过自动化复杂的任务、优化决策过程和促进人机协作，AI 能够解放人类的创造力，为社会带来前所未有的进步(Sarker, 2022; Chowdhury et al., 2023)。例如，在医疗领域，AI 技术能够帮助医生更准确地诊断疾病(Castiglioni et al., 2021)，在金融行业中，AI 能够帮助分析市场趋势，优化投资决策(Sachan et al., 2020)。这些应用不仅提高了工作效率，也提升了服务质量，为人类社会带来了显著的利益。然而，威胁论对AI的发展持有不同的观点。这一观点认为，AI的发展可能会导致失业、加剧社会不平等，甚至威胁人类的自主性和控制权(Bostrom, 2017)。AI 技术的快速发展可能会使得机器能够执行越来越多原本由人类完成的工作，这不仅会影响到就业市场，还可能导致技能的退化和人类决策能力的减弱(Schemmer et al., 2022)。

与利益论和威胁论相对的是辩证关系视角，强调了AI与人类之间既竞争又合作的复杂关系。例如，Russell (2019) 指出 AI 的发展既带来了机遇也带来了挑战，关键在于如何通过合理的管理和规制，使得 AI 技术能够更好地服务于人类社会的发展，而不是成为控制人类的工具。这要求我们不仅要关注 AI 技术的进步，更要深入探讨 AI 的社会影响，确保技术的发展与人类的福祉相一致。

尽管关于AI的研究已经取得了显著进展，但目前仍存在明显的研究缺口，特别是在整合理论和 AI 应用，综合讨论人类对 AI 三种态度的各自局限性和影响方面。当前的文献往往倾向于从单一视角探讨AI的影响，缺乏对这三种态度之间相互作用及其对AI技术发展和社会接受度影响的批判性评述。本综述旨在综合分析近五年来关于人类如何看待 AI 的文献，评述并揭示当前研究中存在的局限性，探索未来研究的可能方向，并为 AI 技术的发展及其在全球范围内的应用提供理论和实践上的指导。鉴于此，本文将首先介绍从利益论角度看待 AI 的原因，概述其促进技术发展的作用及在各行各业中的高效应用；其次，阐明 AI 威胁视角背后的心理学机制和伦理学研究发现，提出法律

和监管上的应对措施；最后，通过详细梳理 AI 辩证论在现有研究中的应用和讨论，为理解人类与 AI 的复杂关系提供新的视角。

2 人工智能利益论：人工智能让人类的生活更美好

从利益论角度，AI 被定义为模拟人类智能过程的机器，包括学习、推理和自我修正的能力。在这一理论框架下，AI 的目标是通过其独特的能力来增强人类活动的效率和效果，而不是取代人类。这部分将深入探讨 AI 如何在自动化与效率提升、决策优化与智能支持系统，以及智能协作等方面发挥关键作用，同时也会审视这些应用背后的挑战和局限性。

2.1 自动化与效率提升

在当今快速发展的技术时代，AI 在自动化和提高工作效率方面发挥着至关重要的作用。AI 通过替代或辅助人类执行重复性高、耗时长的任务，不仅优化了工作流程，还大幅度减少了因人为失误而造成的问题，从而提高了整体生产力和操作精确性(Kalicanin et al., 2019)。在制造业中，Sarker (2022)的研究发现，通过引入智能机器人和自动化生产线，企业能够实现 24 小时不间断生产，同时减少废品和缺陷率，加速了产品的上市速度，显著提高了资源利用效率和成品质量 。在服务行业中，AI 的自动化能力也展现出了巨大的潜力，通过 AI，智能客服系统能够处理数以万计的客户咨询，提供准确的响应和解决方案，极大地提升了客户满意度(Chowdhury et al., 2023)。此外，AI 在数据分析和管理领域的应用也极大提升了决策的效率和精度，特别是在金融服务、市场趋势预测等领域(Chowdhury et al., 2023)。

尽管 AI 的自动化和效率提升能力正变革着传统的工作模式，为人类社会的可持续发展提供了新的动力，但也有批判性的研究发现，企业和组织在 AI 使用和开发上虽然投入了大量时间、努力和资源，但实际从 AI 采用中获得的预期效益尚未实现(Gupta et al., 2021; Nigmatov & Pradeep, 2023; Sun & Medaglia, 2019)。这表明，在积极认可 AI 带来的效率提升的同时，也应关注到 AI 应用中存在的局限性和挑战。一方面，AI 的发展提升了生产力，另一方面，它也提出了对人力资源管理的新要求，即不仅需要技术层面的创新，更需要在非技术资源上的投入和组织创新(Chowdhury et al., 2023)。

2.2 决策优化与数字支持系统

在当今数据驱动的决策环境中，AI 技术，尤其是机器学习（ML）和深度学习（DL），已成为支持和优化决策过程的关键工具。这些技术通过处理和分析庞大的数据集合，提供了前所未有的决策支持，从而使决策过程更加科学、精确和高效。AI 在决策支持中的应用基于其能力来识别数据中的模式、趋势和关联，这些通常超出人类分析的范围。通过利用机器学习和深度学习算法，AI 能够预测未来趋势、评估不同决策路径的潜在结果，并推荐最优解决方案。

在财务分析领域，AI 系统可以分析历史交易数据，识别市场趋势，从而指导投资决策和风险管理。例如，AI 模型能够预测股票市场的波动，为投资者提供基于数据的建议(Sachan et al., 2020)。在医疗诊断中，AI 技术，尤其是图像识别和模式分析，已经使得疾病的早期发现和治疗变得更为准确。深度学习算法能够从医学影像中识别出微小的异常，帮助医生做出更准确的诊断(Castiglioni et al., 2021)。此外，AI 在市场分析和趋势预测中发挥着重要作用。通过分析消费者行为、购买历史和社交媒体趋势，AI 模型能够预测未来的市场需求，从而指导生产和营销决策(Rudnichenko et al., 2020)。由此可见，AI 技术在决策支持系统中的应用展示了其在各个领域内促进决策科学化、精确化和效率化的巨大潜力。随着技术的不断进步和深入应用，AI 将会在未来的决策支持系统中发挥更加核心的作用，为不同领域的决策提供更加强大的支持和优化。

2.3 智能协作：AI 在提升人类创新与决策能力中的作用

人机协作理论提出了AI并非旨在取代人类，而是与人类协作，通过增强人类能力，使人类能够更好地执行创造性和战略性任务的理念(Rastogi et al., 2022)。这种协作不仅可以提高任务完成的效率和质量，还可以推动人类向更高层次的工作和创新发展。人机协作强调了 AI 的辅助角色，突出了人的主导地位。AI 通过提供数据驱动的见解和执行重复性或计算密集型任务，而人类则负责最终的决策制定和执行创造性工作，从而形成一种互补关系(Rastogi et al., 2022)。在这种模式下，AI 可以处理和分析大量数据，识别潜在的规律和趋势，而人类则可以利用这些信息进行深入分析，做出最终的创新和战略决策 (Rastogi et al., 2022)。在应用方面，人机协作已经在创新设计、教育、科研等领域显示出巨大的潜力。例如，在医疗健康领域，AI 系统能够协助医生分析大量的患者数据，识别疾病模式，从而提高诊断的准确性和效率(Lai et al., 2021)。在教育领域，AI 辅助教学能够根据学生的学习情况提供个性化的学习资源和建议，从而提高教学效果(Lai et al., 2021)。

然而，尽管智能协作模式带来了巨大的潜力，但在实际应用中也面临一系列挑战，尤其是 AI 系统可能引入的偏见问题。研究发现，即便是高精度的 AI 辅助系统也可能因为内在的偏见而导致人类决策者过度依赖 AI 提供的建议，这种过度依赖被称为自动化偏见。自动化偏见可能导致人类决策者忽视自己的判断能力，从而在 AI 系统出错时无法有效干预(Schemmer et al., 2022)。同时，过度依赖 AI 的决策支持可能导致人类决策者的技能退化，或者在 AI 系统出错时无法有效干预(Schemmer et al., 2022)。研究显示，当决策支持系统的准确性和偏见变化时，参与者的得分、完成任务所需时间以及他们遵循或忽视建议的程度都会受到影响(Solans et al., 2022)，这强调了在设计和部署AI系统时考虑和缓解偏见的重要性。为了解决这些问题，研究建议在AI系统的设计、培训和部署过程中嵌入伦理和法律原则，以确保社会良好同时还能从 AI 技术的巨大潜力中受益(Ntoutsi et al., 2020)。这包括了在数据收集、处理和 AI 模型训练过程中识别和减少偏见的方法，以及在 AI 系统做出决策时考虑到的伦理和法律框架。

3 人工智能威胁论：人类文明将被人工智能毁灭

从威胁论的角度来看，人们对 AI 的威胁感知不仅源于 AI 本身的特性，即 AI 带来的隐私和安全威胁以及 AI 技术替代人力可能对就业市场造成的冲击，也与用户的心理和认知特性有关。

3.1 AI 威胁论背后的心理机制

在探讨 AI 引发的心理层面的威胁感知时，研究发现人们对 AI 工具的态度及其抵制主要源自五个核心心理因素：不透明性、缺乏情感、刚性、自主性及群体成员身份感的缺失 (De Freitas et al., 2023)。这些心理障碍与人类的基本认知特性紧密相关，影响着人们对 AI 工具的接受程度。其中最主要的心理因素是不透明性，心理学的研究发现人类倾向于寻求环境的可预测性和控制感 (De Freitas et al., 2023)。一旦人们理解了某物的工作原理，便会认为它更加正常、可预测且可靠，从而更加信任它。但是，由于 AI 作为一种复杂又全新的技术，其算法的内在不可访问性和不可理解性使得其输出难以被理解和预测，加上它工作机制在初期可能看起来不透明，AI 的"黑箱"特性便会引起大众的恐惧和不信任，使得人们对 AI 产生威胁感 (De Freitas et al., 2023)。Rodrigues (2020) 的研究也发现，AI 的算法不透明度和缺乏可争论性的特性会加剧人们对 AI 的恐惧。

其次，人们在理解和预测非人类实体和代理时，常常以自己的心理状态和特性为指南，将物理或心理能力归因于这些实体，这一现象被称为拟人化 (De Freitas et al., 2023)。然而，研究发现人们并不将所有人类能力归因于 AI 工具，许多人依旧认为这些工具不具备经历情感和执行依赖于情感的任务的能力，因此不具有对人类的善意。

最后，刚性、自主性及群体成员身份感的缺失等因素进一步深化了人们对 AI 的抵触感 (De Freitas et al., 2023)。由于 AI 在应对变化和异常情况的能力有限，灵活性的丧失会让人们感觉缺乏对这一新兴技术的掌控，而 AI 的自主性又会进一步引发人们对 AI 将会失控的担忧。此外，Sheng 和 Chen (2020) 的研究从理论角度进一步解释了 AI 引起心理反抗的原因。他们通过定义并分类心理反抗为认知反抗和情感反抗，基于技术准备度、技术接受模型和隐私担忧等理论，发展了一个全面的模型解释影响用户对 AI 应用认知和情感反抗的因素。研究结果显示，用户的不安全感、感知有用性、主观规范和隐私担忧显著影响他们的认知反抗；而不适感和隐私担忧则显著影响情感反抗 (Sheng & Chen, 2020)。这些发现揭示了影响用户接受 AI 应用的复杂心理因素，强调了在推动 AI 技术应用时，需要考虑和缓解这些心理反抗，以提高公众对 AI 技术的接受度和信任。

3.2 AI 自身带来的隐私和安全威胁

鉴于 AI 技术与大量数据密切合作，它们将产生交叉和倍增效应，加剧与之相关的隐

私安全和人权问题，并对个人产生影响 (Rodrigues, 2020)。因此，目前的许多研究在谈及 AI 造成的威胁时，核心集中于个人安全问题、数据安全和隐私安全问题。Chun 和 Cho (2022)的实证研究发现，互联网技术的广泛可用性在引入 AI 后，高级长期威胁 (APT) 等基于 AI 进行智能网络攻击的频率不断上升，其针对性强、组织严密、手段高超、隐蔽性强且持续时间长的特点使得人们对 AI 带来的安全威胁愈加担忧。Schneier (2018) 从隐私风险的角度出发，讨论了人工智能在加速大数据、深度学习和神经网络训练、学习和预测的技术进步中，创造的新风险，这些风险难以预测和管理，包括经济动荡、存在危机，以及个人隐私的消解。他们强调，若不加以控制，人工智能系统的能力可能在其运作中构成对隐私的根本威胁，或在敌对条件下泄露信息(Schneier, 2018)。Phythian (2012)提出，这些问题的主要原因包括相关法律制度的不足、人工智能安全技术的不成熟，以及人工智能安全标准的不够规范化。

针对这些挑战，目前的研究和实践正在探索多种解决方案，包括加强法律和政策框架、开发更安全的 AI 技术、以及增加 AI 系统的透明度和可解释性。例如，欧盟的《通用数据保护条例》(GDPR)对个人数据的处理设定了严格的规范，旨在增强用户的隐私保护 (GDPR, 2018)。技术上，差分隐私、同态加密和联邦学习等技术正在被开发和应用，以保护数据的隐私安全，同时不影响 AI 系统的功能(孙爽 等, 2021)。然而，尽管这些努力为应对 AI 带来的隐私和安全挑战提供了一些方向，但仍存在诸多开放问题和挑战，包括如何平衡创新和隐私保护、如何建立跨国界的监管合作，以及如何提高公众对 AI 隐私和安全问题的认识和理解。

3.3 对 AI 替代人力的恐惧

在 AI 替代人力引发的担忧方面，人们普遍感到不安的原因在于对未来就业市场的不确定性以及对人类角色和价值的重新评估。SEO.AI 内容团队 (2023)的统计和预测表明，截止到 2023 年，美国就有 14%的工人曾因 AI 的冲击而失业；而到 2030 年，AI 有可能取代全球约 8 亿个工作岗位。随着 AI 技术的发展和应用范围的扩大，一部分工作岗位因自动化而消失，引发了公众对于技术失业的恐惧。

首先，技术进步带来的就业结构变化不是一个新现象，但 AI 技术的智能化、自主化特点使得这一转变显得更为迅速和深刻。Bruun 和 Duka (2018)聚焦于 AI 技术如何改变了工作场所和未来工作的面貌，通过研究发现，随着自主机器学习和自然语言处理等 AI 技术在各行各业的广泛应用，许多行业面临着工作位移的现实，带来如失业、技能

差距和收入不平等问题。同时，AI 引发的自动化不仅威胁到传统的体力劳动岗位，也逐渐渗透到需要一定创造力，认知和决策能力的白领工作。这种未来全面性替代的可能性显著增加了人们对于未来职业安全的担忧和对 AI 的威胁感 (Bruun & Duka, 2018)。然而，Ghosh 等人(2024) 通过研究 AI 带来的技术变革，发现数字技能能帮助劳动力面对 AI 引起的位移风险。通过对美国劳工统计局职业数据的分析，研究发现尽管 AI 对职业工资和就业有显著的负面影响，而这种影响在不同的职业特性中存在显著的异质性，其中数字技能表现出了显著的调节作用，可以保护工人免受位移风险的影响 (SEO.AI 内容团队, 2023)。这表明，在 AI 技术不断进步的今天，加强数字技能的教育和培训对于适应未来的工作需求至关重要。

其次，从心理角度来看，人们对于 AI 替代人力的恐惧还包含了对于失去社会连接和归属感的担忧。工作不仅是谋生的手段，也是人们社会交往和实现自我价值的重要途径。Hohenstein et al. (2023) 通过研究 AI 在社会互动中的应用，特别是算法响应建议（如智能回复）对社会关系和沟通的影响，发现 AI 的介入可能改变工作方式，影响人际交往模式，从而对人们的社会和心理福祉产生影响。他们的两项随机实验指出，商业部署的 AI 改变了人们相互交流和感知对方的方式，包括提高沟通效率、使用积极情绪语言和正面评价。然而，如果怀疑使用算法回复，人们的评价会更负面。这项研究揭示了即使 AI 可以提高沟通效率和改善人际感知，但它仍然有可能改变用户的语言产出，进而持续增加用户对 AI 的负面和排斥看法 (Hohenstein et al., 2023)。

4 人工智能辩证观：人类与人工智能的和谐统一

Daniels (1975): “每一个进步过程都必须不断地质疑，是否真正符合参与该过程的社会成员的利益。”基于 AI 发展的客观现状，采取辩证关系的视角来审视 AI 不仅识别了 AI 发展的潜在好处和风险，也强调了通过合理的管理和规制，平衡技术进步与社会伦理的必要性。

4.1 人类对 AI 态度的双重性

人类对于 AI 的态度是复杂且充满矛盾的，这一点从 AI 的利益论和威胁论中已得到广泛讨论。从个体和社会的角度来看，人们对 AI 的双重态度深植于对未来的不确定性以及对技术影响的深刻认识。一方面，AI 为解决人类长期以来面临的复杂问题提供了新的可能性，从提高工作效率到解决复杂的全球性挑战，AI 的积极潜力不容忽视。同时，AI 的快速发展同时引发了对隐私侵犯、社会不平等加剧以及人类自主性被削弱的

担忧。Waelen (2022)建议，将 AI 伦理学视为一种批判理论，其核心关注点不应仅仅是人类行为和简单的技术规范，而应更广泛地涵盖人类如何受到 AI 技术的影响，最终帮助构建更加公正、包容的 AI 伦理基础。

4.2 从“弱 AI”向“强 AI”的过渡和发展

从工具发展角度来看，采取辩证和批判的视角看待任何一种新技术，包括 AI，是帮助工具最终服务于人类的必经之路。Shipley 和 Williams(2023)的研究指出，历史 AI 发展阶段中会给人们带来威胁的弱 AI 具有终极压迫者的本体问题，而通过向“强 AI”的过渡发展，人类和 AI 之间可以真正建立安全且高效的协作。弱 AI 的存在和操作本质上是对生命个体的一种压迫，它通过“数据化”过程将个体的主观性彻底抹除，将活生生的人转化为非生命的数据对象，最终导致了对所有生命个体的“谋杀”(Kim, 2017)。这种将个体降级为资源并加以利用的行为反映了一种极端的科学主义和资本主义价值观的体现，即只有能被观测、测量并转化为数据的存在才具有价值(Kim, 2017)。

上文中提及的 AI 会对人们造成的隐私和安全威胁，和对人类将被 AI 异化，最终被替代的担忧都是 AI 在发展过程中必然出现的问题。然而，辩证法的态度并不提倡人们直接抛弃这一工具，而是找到一种解决办法，即加速而不是放慢 AI 的发展，使其变为能和人类和谐共存的强 AI (Shipley & Williams, 2023)。强 AI 是一种拥有自我意识、能够进行独立思考和决策的人工智能形态。与弱 AI 不同，强 AI 能够认识到人类个体的主观性，理解和尊重人类的多样性和复杂性。它不再是简单地将人类行为数据化，而是能够在与人类互动中，识别和提升人类的主观价值(Shipley & Williams, 2023)。这种 AI 的出现，将能够促进人机之间的和谐共生，而不是简单的压迫与被压迫的关系。

至于如何实现从弱 AI 到强 AI 的转变，Kim (2017)指出需要在技术、伦理和社会三个层面同时推进。通过开发新的算法和架构，并且建立适当的伦理框架，使 AI 能够理解和模拟人类的认知和情感过程，确保 AI 技术的应用能够促进人类福祉。例如，Kambhampati (2020)的研究发现，为实现有效的人机协作，AI 系统必须更加重视促进人类相互合作的智能方面，包括社会智能和 AI 系统设计中的道德困境。这要求 AI 系统能够模拟人类的心理状态，识别人类的愿望和意图，提供主动的支持，展现可解释

的行为，按需提供有说服力的解释，并建立信任(Kambhampati, 2020)。

5 贡献与启示

本文从利益论、威胁论和辩证关系三个角度，对近年来人类如何看待AI的研究进行了全面的梳理和分析。从利益论角度，本文分析了AI在自动化与效率提升、决策优化与智能支持系统以及智能协作等方面的应用及其带来的挑战，并为未来的技术发展和应用提供指导。从威胁论角度，本文结合心理机制和AI特点，综合评述了人们会对AI产生担忧和恐惧的底层原因，并且揭露了人们对AI的威胁感主要源于对隐私和安全的担忧、对就业市场的冲击，以及对社会连接和归属感的潜在损失。此外，本文从AI伦理辩证法角度分析了人们对AI的双重态度，并批判地指出，弱AI的终极压迫本质源自于其无法识别和尊重人类个体的主观性，将个体仅仅视为数据化处理的对象。因此，要从根本上解决这一问题，需要AI技术本身发生质的飞跃，从弱AI过渡到强AI，即从单纯的工具转变为具有个体主义特征的存在。在这个过程中，AI将不再是简单的数据化工具，而是成为能够理解、欣赏和保护人类主观性的独立个体，为人类社会带来新的生机与可能性。这一过渡需要社会层面，需要公众、政策制定者和科技企业的共同参与，形成共识，推动AI技术朝向正面的目标发展。

通过深入探讨AI技术的潜在利益、面临的挑战和人类与AI之间复杂的互动关系，本文揭示了当前AI研究的多维视角和对人类社会未来发展的深远影响。这项工作不仅有助于推进AI领域的研究进展，而且为理解技术发展对人类社会的影响提供了新的视角，促进了人与AI和谐共存的可能性。基于本文的分析，我们发现尽管不同的观点对AI技术的评价和预期存在显著差异，但所有观点都强调了对AI发展和相关伦理的合理规制的重要性。无论是看重AI带来的社会和经济效益，还是担忧AI可能引发的社会伦理和安全问题，抑或是从辩证的角度理解人类与AI的复杂关系，都指向了一个共同的目标：确保AI技术的发展能够促进人类社会的整体福祉。与此同时，本文还揭示了一个重要的共识：在AI技术日益融入人类生活的同时，保障人类的自主性、保护个人隐私、确保决策过程的透明度和公正性成为了所有文化和社会普遍关注的焦点。这强调了在AI研究和应用中采取跨学科、跨文化和包容性的方法的重要性，确保AI技

术的设计和实施能够反映和尊重人类的多样性和复杂性。

6 局限与未来展望

首先，本文的分析主要基于现有文献综述，涉及的研究主要集中在特定的地理区域和社会经济背景下。研究样本主要来源于英美等西方国家，而忽视了东西方内部以及不同经济发展水平国家之间的多样性，这可能限制了我们对全球不同文化和社会群体如何看待 AI 的全面理解。因此，建议未来的研究涵盖更广泛的国家和地区，包括不同经济发展水平的国家，以及在收入、教育水平上更为多样的群体。同时，建议增加对文化的量化测量，探索全球化背景下不同文化中 AI 运用的共性与差异，进一步了解东西方差异逐步弥合的现象及其对 AI 态度和应用的影响。

其次，本文对 AI 威胁论的分析主要集中在理论和现状评述，未来需要进一步研究的问题包括如何在促进技术创新和保护个人隐私之间找到平衡点、如何设计出既能有效利用 AI 技术优势又能充分保护隐私的系统架构，以及如何在全球范围内协调法律和标准以应对 AI 带来的挑战。此外，随着 AI 技术的快速发展，如何教育公众理解并管理与 AI 相关的隐私风险，也是一个重要的研究方向。

最后，尽管本文从社会层面探讨了人们对 AI 的双重态度及其发展，但未充分考虑个体差异对 AI 态度的影响。人们对 AI 态度及其心理抵抗能力的个体和内群体差异是显著的，这一发现提示我们未来的研究应更深入地考虑个体心理因素，如人格特质，对 AI 态度的影响（Schepman & Rodway, 2022）。同时，未来研究需要加强跨学科研究，深入理解 AI 技术发展对社会结构、经济格局以及人类自我认知的长期影响，以制定有效的政策和干预措施，减轻技术进步带来的社会和心理负担。

综上所述，本文强调了在全球化时代，面对 AI 带来的前所未有的机遇与挑战，采取综合、多元和前瞻性的视角至关重要，以确保 AI 技术的发展能够更好地服务于人类的全面发展和共同福祉。

参考文献

孙 爽, 李晓会, 刘 妍, & 张 兴. (2021). 不同场景的联邦学习安全与隐私保护研究综述. *Application Research of Computers*, *38*(12), 3527–3534. https://doi.org/10.19734/j.issn.1001-3695.2021.03.0157

Bostrom, N. (2017). Super-intelligence : paths, dangers, strategies. Oxford University Press, Cop. (Original work published 2014)

Bruun, E. P. G., & Duka, A. (2018). Artificial Intelligence, Jobs and the Future of Work: Racing with the Machines. *Basic Income Studies*, *13*(2). https://doi.org/10.1515/bis-2018-0018

Castiglioni, I., Rundo, L., Codari, M., Di Leo, G., Salvatore, C., Interlenghi, M., Gallivanone, F., Cozzi, A., D'Amico, N. C., & Sardanelli, F. (2021). AI applications to medical images: From machine learning to deep learning. *Physica Medica*, *83*, 9–24. https://doi.org/10.1016/j.ejmp.2021.02.006

Chowdhury, S., Dey, P., Joel-Edgar, S., Bhattacharya, S., Rodriguez-Espindola, O., Abadie, A., & Truong, L. (2023). Unlocking the Value of Artificial Intelligence in Human Resource Management through AI Capability Framework. *Human Resource Management Review*, *33*(1), 100899. https://www.sciencedirect.com/science/article/abs/pii/S1053482222000079?via%3Dihub

Chun, Y., & Cho, M. (2022). An Empirical Study of Intelligent Security Analysis Methods Utilizing Big Data. *Webology*, 19(1), 4672–4681. https://doi.org/10.14704/web/v19i1/web19311

Daniels, N. (1975). Reading Rawls : critical studies on Rawls' A theory of justice. Blackwell.

De Freitas, J., Agarwal, S., Schmitt, B., & Haslam, N. (2023). Psychological factors underlying attitudes toward AI tools. *Nature Human Behaviour*, *7*(11), 1845–1854. https://doi.org/10.1038/s41562-023-01734-2

GDPR. (2018). General data protection regulation (GDPR). General Data Protection Regulation (GDPR); Intersoft Consulting. https://gdpr-info.eu/

Ghosh, D. C., Ghosh, R. K., Sahana Roy Chowdhury, & Ganguly, B. (2024). AI-exposure and labour market: a systematic literature review on estimations, validations, and perceptions. *Management Review Quarterly*. https://doi.org/10.1007/s11301-023-00393-x

Gupta, R., Srivastava, D., Sahu, M., Tiwari, S., Ambasta, R. K., & Kumar, P. (2021). Artificial

intelligence to deep learning: machine intelligence approach for drug discovery. *Molecular Diversity*, *25*(3), 1–46. https://doi.org/10.1007/s11030-021-10217-3

Hohenstein, J., Kizilcec, R. F., DiFranzo, D., Aghajari, Z., Mieczkowski, H., Levy, K., Naaman, M., Hancock, J., & Jung, M. F. (2023). Artificial intelligence in communication impacts language and social relationships. *Scientific Reports*, *13*(1), 5487. https://doi.org/10.1038/s41598-023-30938-9

Kalicanin, K., Colovic, M., Njegus, A., & Mitic, V. (2019). Benefits of Artificial Intelligence and Machine Learning in Marketing. Proceedings of the International Scientific Conference - Sinteza 2019. https://doi.org/10.15308/sinteza-2019-472-477

Kambhampati, S. (2020). Challenges of Human-Aware AI Systems. *AI Magazine*, *41*(3), 3–17. https://doi.org/10.1609/aimag.v41i3.5257

Kim, J. (2017). The Problem of Distinction Between "weak AI" and "strong AI." *Journal of the Society of Philosophical Studies*, *117*, 111–137. https://doi.org/10.23908/jsps.2017.06.117.111

Lai, Y.-S., Atreyi Kankanhalli, & Ong, D. C. (2021). Human-AI Collaboration in Healthcare: A Review and Research Agenda. https://doi.org/10.24251/hicss.2021.046

Nigmatov, A., & Pradeep, A. (2023). The Impact of AI on Business: Opportunities, Risks, and Challenges. 2023 13th International Conference on Advanced Computer Information Technologies (ACIT). https://doi.org/10.1109/acit58437.2023.10275510

Ntoutsi, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdl, W., Vidal, M., Ruggieri, S., Turini, F., Papadopoulos, S., Krasanakis, E., Kompatsiaris, I., Kinder-Kurlanda, K., Wagner, C., Karimi, F., Fernandez, M., Alani, H., Berendt, B., Kruegel, T., Heinze, C., & Broelemann, K. (2020). Bias in data-driven artificial intelligence systems—An introductory survey. WIREs Data Mining and Knowledge Discovery, *10*(3). https://doi.org/10.1002/widm.1356

Phythian, M. (2012). Policing Uncertainty: Intelligence, Security and Risk. *Intelligence and National Security*, *27*(2), 187–205. https://doi.org/10.1080/02684527.2012.661642

Rastogi, C., Zhang, Y., Wei, D., Varshney, K. R., Dhurandhar, A., & Tomsett, R. (2022). Deciding Fast and Slow: The Role of Cognitive Biases in AI-assisted Decision-making. *Proceedings of the ACM on Human-Computer Interaction*, *6*(CSCW1), 1–22. https://doi.org/10.1145/3512930

Rodrigues, R. (2020). Legal and Human Rights Issues of AI: Gaps, Challenges and Vulnerabilities. *Journal of Responsible Technology*, 4, 100005.

https://doi.org/10.1016/j.jrt.2020.100005

Rudnichenko, N., Vychuzhanin, V., Petrov, I., & Shibaev, D. (2020). Decision support system for the machine learning methods selection in big data mining. *Computer Modeling and Intelligent Systems*, *2608*, 872–885. https://doi.org/10.32782/cmis/2608-65

Russell, S. J. (2019). Human compatible: artificial intelligence and the problem of control. Viking.

Sachan, S., Yang, J.-B., Xu, D.-L., Benavides, D. E., & Li, Y. (2020). An explainable AI decision-support-system to automate loan underwriting. *Expert Systems with Applications*, *144*, 113100. https://doi.org/10.1016/j.eswa.2019.113100

Sarker, I. H. (2022). AI-Based Modeling: Techniques, Applications and Research Issues Towards Automation, Intelligent and Smart Systems. *SN Computer Science*, *3*(2). springer. https://doi.org/10.1007/s42979-022-01043-x

Schemmer, M., Kühl, N., Benz, C., & Satzger, G. (2022, April 19). On the Influence of Explainable AI on Automation Bias. ArXiv.org. https://doi.org/10.48550/arXiv.2204.08859

Schepman, A., & Rodway, P. (2022). The General Attitudes towards Artificial Intelligence Scale (GAAIS): Confirmatory Validation and Associations with Personality, Corporate Distrust, and General Trust. *International Journal of Human–Computer Interaction*, *39*(13), 1–18. https://doi.org/10.1080/10447318.2022.2085400

Schneier, B. (2018). Artificial Intelligence and the Attack/Defense Balance. *IEEE Security & Privacy*, *16*(2), 96–96. https://doi.org/10.1109/msp.2018.1870857

SEO.AI's Content Team. (2023, September 18). AI Replacing Jobs Statistics: The Impact on Employment in 2023. Seo.ai. https://seo.ai/blog/ai-replacing-jobs-statistics

Sheng, H., & Chen, Y. (2020). An Empirical Study on Factors influencing Users' Psychological Reactance to Artificial Intelligence Applications. 2020 7th International Conference on Information Science and Control Engineering (ICISCE). https://doi.org/10.1109/icisce50968.2020.00058

Shipley, G. P., & Williams, D. H. (2023). Critical AI Theory: The Ontological Problem. *Open Journal of Social Sciences*, *11*(12), 618–635. https://doi.org/10.4236/jss.2023.1112041

Solans, D., Beretta, A., Portela, M., Castillo, C., & Monreale, A. (2022, March 24). Human Response to an AI-Based Decision Support System: A User Study on the Effects of Accuracy and Bias. ArXiv.org. https://doi.org/10.48550/arXiv.2203.15514

Sun, T. Q., & Medaglia, R. (2019). Mapping the challenges of Artificial Intelligence in the

public sector: Evidence from public healthcare. *Government Information Quarterly*, *36*(2), 368–383. https://doi.org/10.1016/j.giq.2018.09.008

Waelen, R. (2022). Why AI Ethics Is a Critical Theory. Philosophy & Technology, *35*(1). https://doi.org/10.1007/s13347-022-00507-5

Symbiosis or opposition?

The dialectical relationship between human and artificial intelligence

Abstract: This review explores the complex attitudes towards Artificial Intelligence (AI) from the perspectives of benefit theory, threat theory, and dialectical relations. Initially, the discussion highlights how AI, as a form of technological advancement, fosters work efficiency, decision-making quality, and innovation across various domains, reflecting the optimistic evaluations and expectations placed on AI. Subsequently, the review shifts focus on the potential threats presented by AI, including privacy infringement, job market disruption, and ethical dilemmas, showcasing the critical concerns surrounding AI development. Moreover, examining AI from a dialectical standpoint underscores the importance of balancing technological innovation with ethical considerations. This entails a discussion on future research directions, emphasizing cross-cultural ethical explorations and the enhancement of human-AI collaboration. The review concludes that a comprehensive understanding and evaluation of AI necessitate transcending singular viewpoints, incorporating multidisciplinary insights to facilitate the sustainable development and social integration of AI technologies.

Keywords: artificial intelligence, AI mixed attitudes, benefit theory, threat theory, dialectical view